DOCUMENT RESUME

ED 133 372 TM 005 990

AUTHOR Powell, J. C.
TITLE Evidence for a Phase and Stage Developmental Sequence Derived from Response Patterns on Multiple Choice Tests.
PUB DATE [Sep 76]
NOTE 69p.; Paper presented at the Annual Meeting of the American Psychological Association (Washington, D.C., September 7, 1976) ; Best copy available; Some pages will be marginally legible
AVAILABLE FROM J.C. Powell, Faculty of Education, University of Windsor, 600 Third Concession Road, Windsor, Ontario, Canada, N9B-3P4 ($2.00)

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Behavior Change; Developmental Stages; *Divergent Thinking; *Multiple Choice Tests; *Response Style (Tests); Statistical Analysis; *Testing Problems
IDENTIFIERS Test Theory; *Wrong Answer Analysis

ABSTRACT

The results of five studies into the characteristics of wrong answers as a class of divergent behavior are presented. The evidence from these studies, when taken in combination, suggests that the tendency of researchers to ignore wrong answers has been a fundamental procedural error of broad scope and serious consequences. Instead of the straight line development commonly found when right answers are considered alone, evidence for a phase and stage sequence was found. These results contradict the use of linear models to describe development. Implications to education, to research procedures, to test theory and analysis, and to learning theory are drawn. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED133372

BEST COPY AVAILABLE

EVIDENCE FOR A PHASE AND STAGE DEVELOPMENTAL SEQUENCE DERIVED FROM RESPONSE PATTERNS ON MULTIPLE CHOICE TESTS

by

J. C. Powell

Faculty of Education

University of Windsor

Windsor, Ontario
Canada

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICRO FICHE ONLY HAS BEEN GRANTED BY

J. C. Powell

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

U S DEPARTMENT OF HEALTH EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM005 990

EVIDENCE FOR A PHASE AND STAGE DEVELOPMENTAL SEQUENCE DERIVED FROM RESPONSE PATTERNS ON MULTIPLE CHOICE TESTS

by

J. C. Powell

Faculty of Education

University of Windsor

Windsor, Ontario
Canada

This is a copy of a paper presented to the Annual convention of the American Psychological Association, Washington D. C. September 7, 1976.

© Copyright, J. C. Powell, 1976.

ABSTRACT

The results of 5 studies into the characteristics of wrong answers as a class of divergent behavior are presented. The evidence from these studies, when taken in combination, suggests that the tendency of researchers to ignore wrong answers has been a fundamental procedural error of broad scope and serious consequences. Instead of the straight line development commonly found when right answers are considered alone, evidence for a phase and stage sequence was found. These results contradict the use of linear models to describe development. Implications to education, to research procedures, to test theory and analysis, and to learning theory are drawn.

ABRIDGED FROM THE PROVERBS TEST

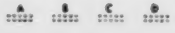
Best Answer Form

Donald R. Gorham, Ph.D.

Directions: Below are some proverbs and you are supposed to indicate what they mean. You are to black in below the letter on the Answer Sheet which is the same as the best answer to each proverb; the one which best explains what the proverb means. Here are two samples:

A. DON'T CROSS THE BRIDGE UNTIL YOU GET TO IT. A.

a. The bridge is a long ways off.
b. People won't like you if you are cross.
c. Don't worry about troubles until they come.
d. Don't be foolish.

B. DON'T CRY OVER SPILT MILK. B. 

a. It won't do any good to cry.
b. Don't be concerned about mistakes of the past.
c. Stop crying and clean it up.
d. It is better to laugh than to cry.

© Psychological Test Specialists 1956

REPRODUCED WITH PERMISSION.

1. RICHES SERVE A WISE MAN BUT COMMAND A FOOL.
   a. Don't let money go to your head.
   b. The poor work for the rich.
   c. Money may help or hinder, according to the individual.
   d. Don't beg, borrow or steal.

2. THE MORE COST, THE MORE HONOR.
   a. For honor and society, it costs.
   b. The harder a thing is to get, the more you appreciate it.
   c. The higher the price, the better a thing is.
   d. Good things have to be paid for in some way.

3. GOLD GOES IN AT ANY GATE EXCEPT HEAVEN'S.
   a. No one can be as good as gold.
   b. Anyone would take money.
   c. Fortune only comes to those who work for it.
   d. You can't buy morals.

4. THERE'S MANY A SLIP TWIXT (BETWEEN) THE CUP AND THE LIP.
   a. Something can happen at the last minute.
   b. Don't talk too much while eating.
   c. A lot can happen between plan and completion.
   d. Don't talk about people too much.

5. ALL IS NOT GOLD THAT GLITTERS.
   a. Don't let temptation get you.
   b. Other things than gold glitter, too.
   c. Everything that looks good isn't necessarily good.
   d. Some things may fool you.

6. SPEECH IS THE PICTURE OF THE MIND.
   a. To have good speech will always help you.
   b. Words paint pictures in the mind.
   c. Speech can accomplish a lot of things.
   d. You are judged by what you say.

7. DON'T THROW GOOD MONEY AFTER BAD.
   a. Don't gamble with a cheater.
   b. Be wise and think of the future.
   c. When you've lost out in something, accept the fact.
   d. Don't waste your money.

8. THE HOT COAL BURNS, THE COLD ONE BLACKENS.
   a. Impetuous action may hurt your reputation.
   b. The burned child avoids the fire.
   c. Extremes of anything are bad.
   d. Leave dangerous things alone.

EVIDENCE FOR A PHASE AND STAGE DEVELOPMENTAL SEQUENCE DERIVED FROM RESPONSE PATTERNS ON MULTIPLE CHOICE TESTS

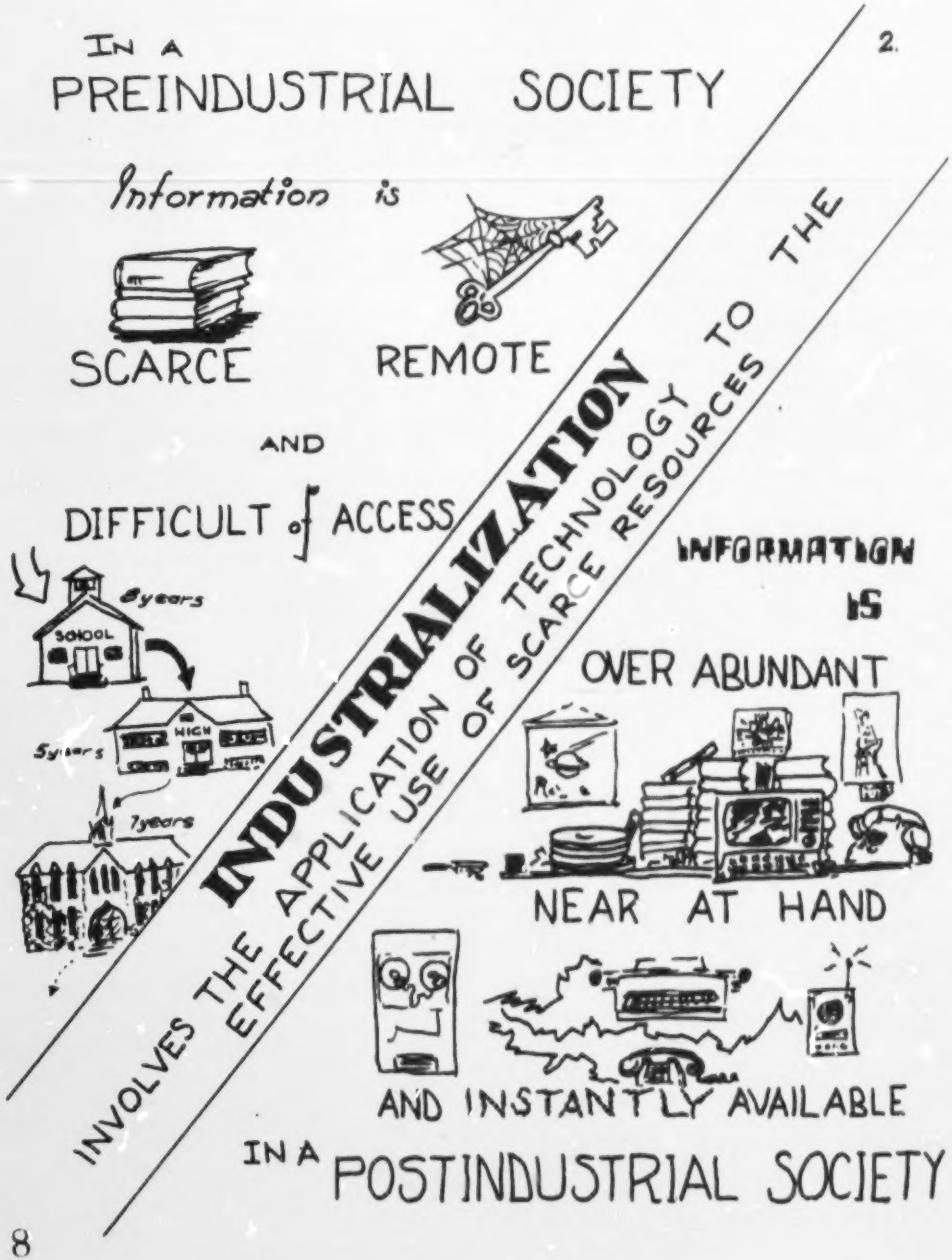
INTRODUCTION

The educational enterprise has been a favorite focus of critics for many years.

There are probably several reasons for this observation. To begin with education is both a conspicuous and a relatively expensive public institution affecting all of our lives at least to some degree almost daily. In this respect education commonly gets a good deal of "bad press" from the media since educators do not typically devote much effort into public relations activities. Many teachers, uneasy about their role in society, openly criticize education as well.

A second reason why education is frequently criticized is that people hold differing views of what education is and what it should do. Carlton (1974) lists and describes seven popular images of the school and Getzels (1974) describes four images of the classroom and visions of the learner. Powell and Cottrell (1976) using a three dimensional adapting of Carlton's approach found eight (8) such images with a residule of about one third of the population not classifiable. In some populations this "not defined" group can be as low as 16 percent (Powell 1976a) suggests that a three dimensional bipolar model is sufficient to describe most populations. With at least 8 views of the role of school in our society, it is not surprising that there is considerable disagreement among various advocates. In all probability, each view has a particular, explicit function, and each would be viable in appropriate contexts.

A third reason for criticism arises from the observation that, in special circumstances, learners sometimes make spectacularly better progress,



at least for short periods of time, than is typical in the general educational setting. The Hawthorne Effect has long been known, however, research evidence on precisely how to establish and to maintain this effect for protracted periods of time has proven to be much less clear.

A fourth reason for some criticism is the belief among certain individuals, that what the school is teaching is not appropriate for the needs of children and/or the future of our society.

In spite of all of these several sources of discomfort, our educational system has, in general, served us well to date. In fact, much, if not most, of our present high level of technological development can be traced, either directly or indirectly to our educational system.

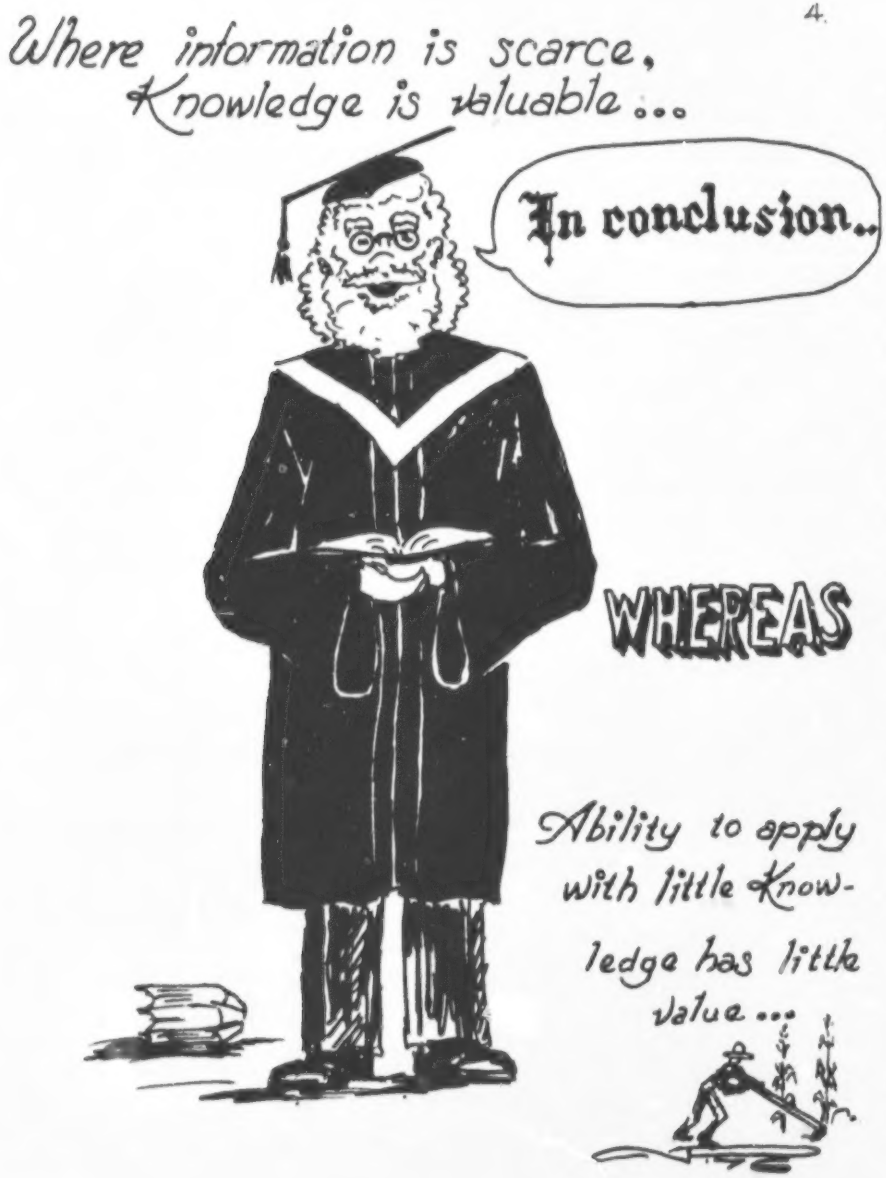
It is within this context that this present paper will discuss a radically different approach to education than is now in common practice. The discussion will provide the basic research evidence which both challenges present common practice and supports the alternative.

a. <u>The needs of a effective educational system</u>

1. <u>For an industrial society</u>.

To begin with, our present educational system has had two main interrelated thrusts that have been the basis of its success. The first has been the transmission of our cultural heritage, and the second has the development if high order generalized intellectual awareness among at least a limited number of citizens. The cultural transmission has enhanced a life style which has focused upon the use of increasingly high order technologies, making an expanding market for these technologies readily available. The development of at least a few individuals of high order intellectual awareness has made it possible to produce the technologies to supply the culturally generated market.

At the beginning of the industrial revolution such information about



our environment as was available was either embedded in the proverbs and moral structures of the uneducated general populace, or was remote and difficult of access. Since information held in memory is remarkably portable as compared with the portability of a library, those few individuals who possessed a high level of education (in the sense of accumulated information) were in great demand. The great need for accurate information brought on by the industrial revolution made the holders of this information extremely valuable.

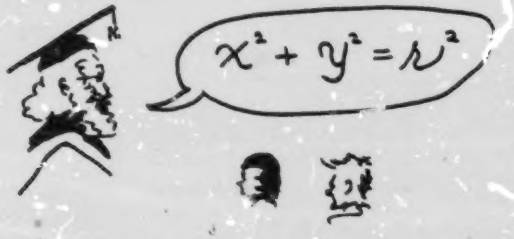
The tradition from which this respect for information arose included an assumption that truth is absolute. That is, knowledge defined in terms of accurate information is unchanging. Hence the educational system which developed was designed specifically to transmit these universal truths in unchanging form to as many people as could successfully absorb them. The entire thrust of education was toward convergent behavior with success measured in terms of how much of what was presented could be accurately returned to the teacher from the learner.

It was this pressure toward convergency which, generated by the demands of various technologies, generated the information transmission heritage of our present schools. In this approach, only those individuals who survive long enough to accumulate sufficient information to be trusted with the responsibility for discovering new information are allowed to develop the additional skills needed for developing new technologies. These skills are usually not engaged in systematically until graduate school. If these advanced students failed to gain these new skills they went into teaching rather than research. Some pursued both teaching and research by basing their activities in a university.

With the demands of the system focused upon convergent behavior, law and order, and the accumulation of unquestioned information, became the two

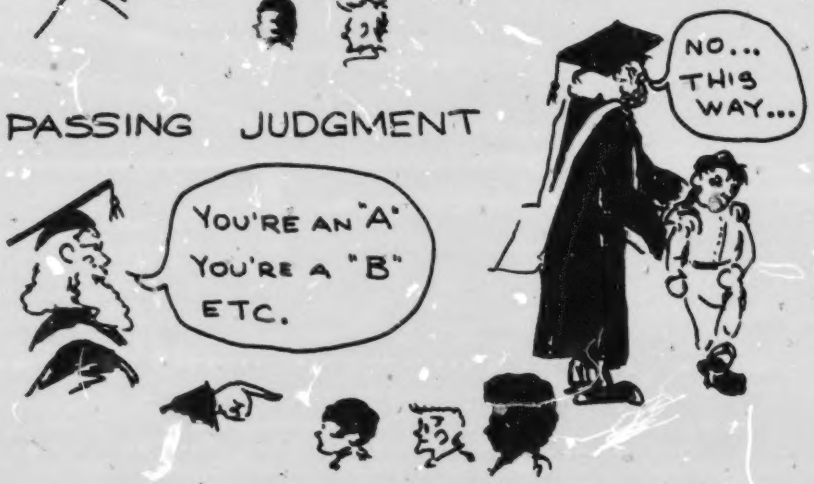
When Knowledge is valuable
Teaching involves...

PRESENTING INFORMATION



PERSUADING

PASSING JUDGMENT



THAT IS - THE 3P'S OF

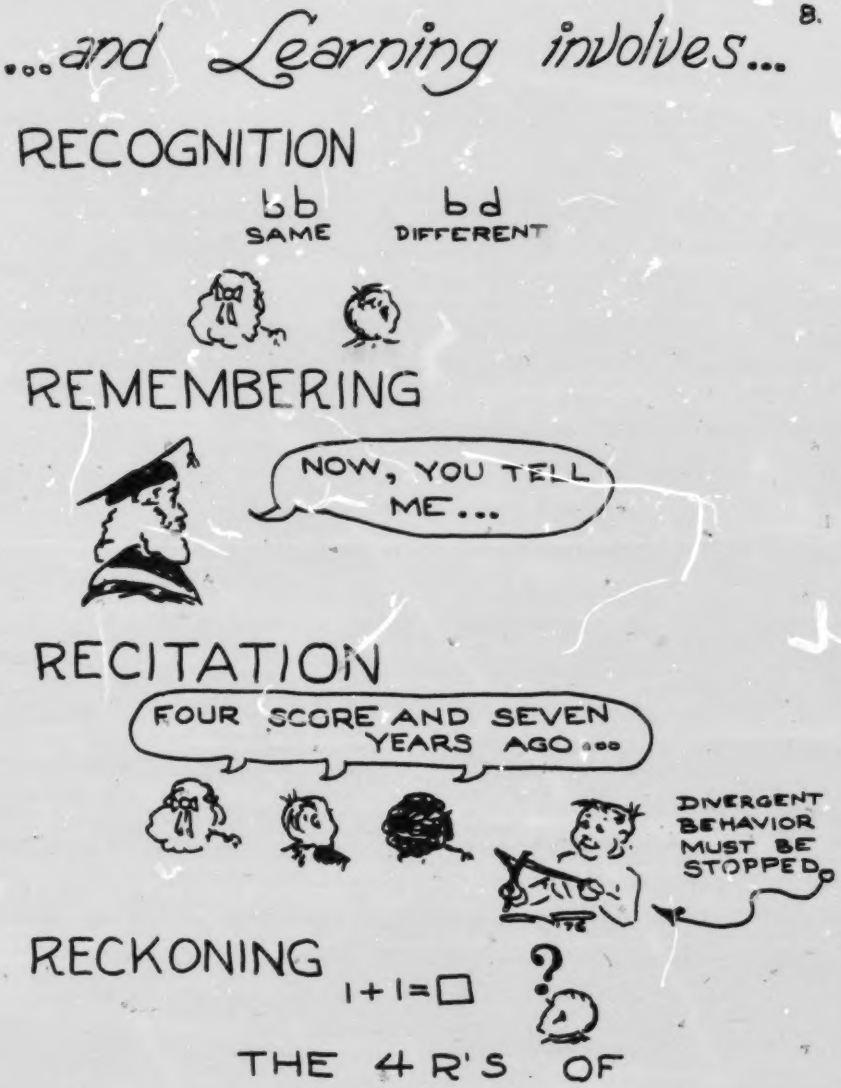
MASS COMMUNICATION

main characteristics of our schools. Teachers presented information, persuaded children to pay attention and to do as they were told, and passed out rewards and punishments in terms of the degree to which each child met these expectations.

The approach to education became essentially one of mass communication. The school resembled an assembly-line. This pattern was not only justified by its success, but also by the success of mass communication and of the assembly line in other segments of our society. In all, the total result was very successful in that it made possible the demand for a high standard of comfort in living currently seen in our society. Also, the differential ability of children to remember unquestioned information out of context provided an admirable statification device. This information was more in context with highly verbal and literate families then with those who did not show these traits. The children of the highly literate had an advantage. Until about 15 years ago the range of intellectual demands required for the employment available very nearly matched the range of educational accomplishment produced by the schools.

It should be made clear, that "thinking" was not neglected in this system. It is only that analysis was stressed over synthesis and creativity. The results of the "thinking" exercises supplied were predictable and converged upon particular generally accepted conclusions. These conclusions are commonly known in the schools as "right answers."

Another important aspect of this approach is the way in which divergent behavior is treated. Since objectives are set in advance and are specific and fixed, departures are treated as deviance to be ignored (negatively reinforced), discouraged or punished (aversively reinforced).



THAT IS- BEHAVIOR MUST CONVERGE TOWARD SPECIFIED TARGETS.

Thus the activities of children have typically centered upon the 4 R's of Behavior Modification. These are 1) Recognition - the ability of the child to discriminate among stimuli and to respond in such a manner that it is clear that s/he has identified the appropriate components of the stimuli in the conventionally accepted manner. 2) Remembering - this is the ability of the child to retain the discriminations taught as a consistent, a continuing and a long term aspect of behavior. 3) Recitation - which refers to the ability of the child to reproduce, upon command, appropriate segments of the originally discriminated stimuli in any specified combination using any of the several media available to him or her. 4) Reckoning - which refers to the ability to select, upon command, a particular data transformation procedure and to use it in such a manner that it produced the expected result. Arithmetic most commonly comes to mind but reckoning is by no means confined to arithmetic.

It is important to note that all of these processes are measurable on the basis of particular behavioral outcomes. That is, success is determined upon the basis of the ability of the learner to converge upon the behavioral outcome defined by the teacher.

Considerable research has been conducted into the factors which produce effective convergence. The key to success in this area has been shown to be contingency management analysis. That is, the critical aspect of the learning environment which must be manipulated in order to produce effective convergence of behavior involves the reward and punishment system established within that environment.

An important aspect of behavior not commonly considered under contingency management procedures is divergent behavior. Of course, divergent behavior is considered by behavior modifiers when this behavior is distructive, making it desireable to eliminate it.

10.

When Information is Over-abundant

Knowledge without application ability is of little value...

The ability to organize and utilize Information is of great value.

2. In a post industrial society.

As successful as this approach has been, there are some fundamental weaknesses to this approach to education. First, any approach to education which is convergent, which tends to stress making everyone alike - or, if you wish - to produce "interchangeable people" is philosophically adhorant to those of us who value individuality.

Second, exclusive attention to convergence is empirically contradictory to the large and growing body of evidence suggesting that individuals differ.

Third, the advent of advanced technology has not only produced a surfiet of information; it has also made it potentially infinitely portable and instantaneously available through electronic data storage, processing, and telecommunication systems.

Finally, research in the hard sciences, particularly since the turn of the century, have made it clear that a set of absolute truths which apply universally cannot be empirically obtained if, indeed, they exist. There is no such thing as unquestionable information. There are always many situational and outcome expectation parameters which must be taken into account before any observation can be interpreted. For this reason, teaching toward a fixed set of expectations may be contradictory to the fundamental nature of the universe.

Where the person who was a repository of knowledge was once of high value, such individuals are now in some degree and in increasing numbers in surplus. On the other hand, individuals who can manage information effectively are in as scarce supply as ever. With the population, pollution, and other problems now increasing in severity, the demand for the effective solution of problems is increasing more rapidly than our supply of high level problem solvers.

The principal skill now in very short supply is effective decision-making under uncertainty. Fulfilling the requirements of supplying an adequate

When Information is abundant...

ADDITIONAL SKILLS ARE NEEDED

THEREFORE

Teaching also involves

DEVELOPING ALTERNATIVES
DISCOVERING NEW RELATIONSHIPS
DISENGAGEMENT FROM OVERT CONTROL

THE 3 D'S OF
FACILITATION...

WHILE *Learning also involves*

INDUSTRY
INITIATIVE
INVESTIGATIVE SKILLS
INNOVATION

THE 4 I'S OF
PARTICIPATORY LEARNING...

THAT IS - BEHAVIOR MUST DIVERGE IN CONSTRUCTIVE DIRECTIONS.

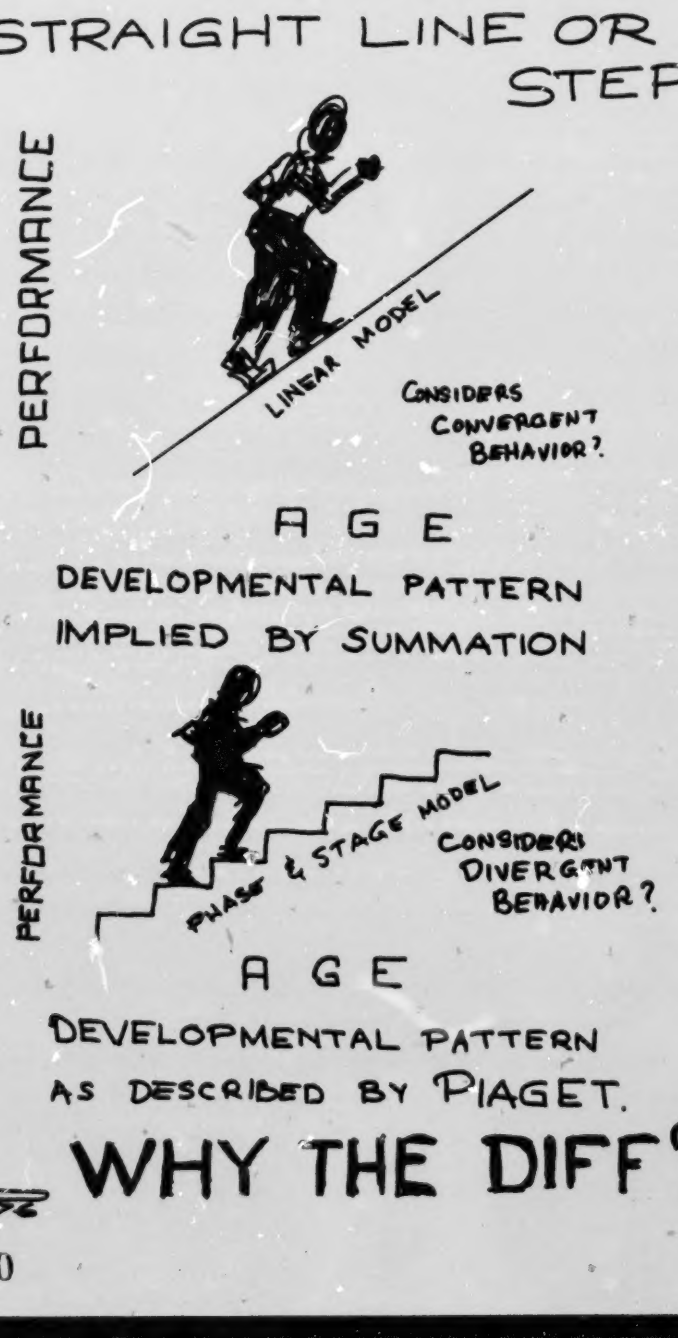
number of people who can deliver these skills may bring about a new "human revolution" as dramatic in scope and as spectacular in consequences as was the industrial revolution before it.

To develop such individuals, who respond with constructive divergence to their environment, will probably require the same fundamental skills as we now develop in our schools -- namely the 4 R's of recognition, remembering, reciting and reckoning. However, such individuals will probably need additional skills relating to the constructive use of divergent behavior patterns.

The three D's of facilitations of 1) developing alternatives 2) discovering new relationship and 3) disengaging from overt control are already well known and have been discussed extensively. Teachers in an "open" setting are enjoyned to create learning situations in which more than one solution is possible. They are encouraged to produce settings in which the children develop or "discover" their own conceptual frame work for particular problems. They are told that motivation is best when it arises from within the individual. The teacher must, therefore, gradually move from overt to covert control systems so that the child can develop self-directing skills.

Attempts to accomplish these objectives have met with mixed success, perhaps because we have been able to tell teachers what to do but not how to do it. It is very difficult to effectively develop divergent behavior in a setting which traditionally supports and enhances convergent behavior. Also, the success of divergent approaches are commonly measured on tests of convergent behavior.

A major change of emphasis in the schools is needed to accommodate such a shift in teacher activities. However, the initial need to accomplish this change of emphasis is some key to the analysis of divergent patterns. How can the child's progress in the 4 I's of participatory learning (Industry, initiative investigative skills, and innovation) be observed if the precise



outcomes of these behaviors cannot be predicted in advance. Without this ability to predict outcomes, our present approach has no means of evaluating progress against a specific set of expectations.

3. Contradictions from methodology.

Typically research evidence in the literature (such as Jamison, Suppes and Wells 1974; and Walker and Schaffarzick - 1974) do not show qualitative and/or differential effects in educational procedures when cumulative (normative) measures are used.

Developmental studies generally find that although specific individuals may differ widely, progress is usually observed to be a sloping straight line which is positively related to age. The great majority of these studies, however, use some convergent criterion for studying the progression under analysis. Most typically in educational studies, these criteria involve total correct scores on educational tests.

In contrast, the writings of Piaget and his co-workers, disclose qualitative differences among children and a phase and stage progression in development. Piaget's approach, however, has been to consider the manners in which children differ (or diverge) from adult behavior and capabilities.

Thus there are different observational conclusions which seem to emerge from different observation procedures. Could it be that these differences are artifacts of the observational procedures used? In this case, which of these observational procedures give the most empirically satisfactory description of the actual transformations which occur as maturation and learning progress? Is there room for optimism that alternative observational strategies can lead to alternative and (hopefully) more effective educational results?

4. Summary of Introduction

The discussion thus far has suggested that much of the current criticism of education may arise from differing perceptions of what education is trying

IN SUMMARY...

In POSTINDUSTRIAL SOCIETY

AN **ADDITIONAL SET** OF

BASIC SKILLS ARE NEEDED BY

Both THE TEACHER AND THE LEARNER.

THESE SKILLS RELATE TO

DIVERGENT MENTATION.

AN IMPORTANT

CATEGORY OF DIVERGENT

BEHAVIOR

WHICH HAS HAD

LITTLE RESEARCH ATTENTION

is

WRONG ANSWERS

to accomplish.

Education has, to date, played an important role in the development of our current high level of technology. It has supported this development by utilizing an effective mass communication procedure successfully accumulating specific knowledge in a convergent environment has built a solid structure of accomplishment.

However, changes that have occurred in the size and availability of the information base upon which our technology is built as a result of that technology have pr duced a fundamental change in the nature of education required. The need has shifted from the ability to correctly perform specific tasks to a new need to be able to make effective decisions and to take prompt appropriate action under conditions of uncertain outcomes.

Put another way the new requirements demand people who have highly developed constructive divergent capabilities. Such individuals are now in very short supply, possibly because of the highly convergent emphasis of our present educational system. This emphasis may tend to discourage divergence.

What is now needed is an approach to observation which will make possible the educational development of divergence. To this end the present paper reports the results of several studies conducted over the past 10 or so years which have focussed upon a particular, class of readily available divergent responses; namely - WRONG ANSWERS.

STATEMENT OF THE PROBLEM

In the study of divergent behavior, wrong answers on multiple choice tests are particularly useful because these are tightly specified and yet occur at the rate of three or four to one convergent (right) answer.

There are, however, several problems in this procedure which must be accommodated.

FIGURE 0.1

BASIC TEST THEORY

1. $R = T + E$

2. $N = R + W + O$

3. $R = \sum_{i=1}^{N} x_i$

IN ENGLISH:

A STUDENT'S TRUE SCORE (T) ON A TEST IS ASSUMED TO BE THE SUM (OR COUNT) $\left(R = \sum_{i=1}^{N} x_i\right)$ OF THE SET OF ANSWERS CONSIDERED TO BE CORRECT (R) ADJUSTED BY AN UNKNOWN MEASUREMENT ERROR (E). THE TOTAL POSSIBLE (N) IS THUS THE SUM OF RIGHT, WRONG & OMITTED.

Within the typical assumptions currently made in the educational testing field studying wrong answers from multiple choice tests is a waste of time. We will consider these assumptions and their implications to the present studies. The basis for these assumptions is summarized in Figure 0.1 (opposite)

1. Assumptions made in test theory

a. The KNOW-GUESS assumption

Test theory begins with the fundamental assumption that there are two classes of behavior to be observed by a test; 1) Right answers and 2) wrong answers. A learner's success is determined by the relative proportion of members in the right answer category. That is, successful learners demonstrate this success by having the bulk of their behavior converge upon the behaviors included in the right answer category.

Implied in this assumption is the proposition that right answers can be ordered into a countable set which can be treated as a relatively uniform homogeneous ordinal or interval scale.

Where wrong answers on multiple choice tests are concerned, these are assumed to be guesses. If guessing responses are blind - chance events then each wrong alternative should have about the same number of the students selecting it. This assumption is the basis for the guessing correction often used on multiple choice tests, since by chance alone some of the right answers will also be guesses.

If this assumption is correct any study of wrong answers is a waste of effort.

b. The LINEAR-DEPENDENCY assumption.

This assumption has several aspects. To begin with, since there are only two categories of behavior (right and/or wrong) it is necessary and sufficient to know how many right answers a person has selected. This concept of necessary and sufficient is a mathematical and a philosophy of science concept.

It is related to the concept of redundancy. Having successfully established a fact once it is not necessary to establish it again. Subsequent replications are redunant. The concept of "necessary and sufficient" is, therefore, an attempt to find the smallest possible reduction of a set of data without loss of information.

If there are only two behavior categories, we need only know how many right since the number not right is the difference between the number right and the number possible. Of course the "not right" set consists of "wrong" and "omitted" answers. Omitted answers tell us little about the learner. There are no meaningful differences among wrong answers under the KNOW-GUESS assumption. Hence the total number correct gives us all the information about an individual which we can obtain from one respondent.

Error analysis in diagnostic testing uses wrong answer information, but in this case total correct scores have relatively little meaning. In criterion referenced tests the categorical function of items shifts to DO - CANNOT DO from KNOW - DON'T KNOW. There are some technical considerations which make a difference in the mathematical procedures used, but the logic is similar.

Finally, even if there is more than one category of wrong answer which is meaningful, the assumption that scores can be counted or added still produces linear dependencies. The use of addition in the mathematical procedures make it impossible to extract more information from the results even if this information is present.

In order to use wrong answers for information, it is necessary to use procedures which by-pass this mathematical problem, otherwise working with wrong answers will gain nothing.

3. Wrong answers contain no USEFUL INFORMATION

It is clear that wrong answers are useful in diagnostic testing.

However, diagnostic tests generally have narrow application and highly specific usefulness. They would seem to have little application with respect to general achievement testing.

It is difficult to conceive what information wrong answers might contain. What is more useful than knowing how well a student did on a test?

4. Implications of refutation

On the other hand, if all three of these assumptions can be proven to be false, then serious implications to educational practice emerge.

If the total correct score on a test is neither a necessary nor a sufficient set (nor both) to provide all the useful education information contained in the answers on that test (without redundancy) then the common practice of using such scores may be a fundamental procedural error. Such a demonstration would reconcile these observations with Piaget's findings by demonstrating that the straight line "observation" was spurious.

An additional implication, should refutation occur, is that much of the results of educational research to date would be rendered invalid because of the use of inappropriate mathematical procedures.

In much well conceived and executed research, the key to the lock with respect to convergent behaviors has been clearly shown to be contingency management analysis. Furthermore most of this research has used frequency counts of well defined specific behaviors as its basic data. Such data forms a ratio scale and in this case rectilinear mathematical models are quite appropriate.

It is the much less well defined category of convergent behavior of "right answers" which is the concern here.

Now that constructive divergent behavior has increased so much in importance, we may regard the fostering of divergent behavior as a second educational lock looking for a key.

THE STARTING POINT:

FIGURE 0.2

OBSERVATION 1

EXPECTED DISTRIBUTION OF RESPONSES

17	49	17	17

RIGHT ANSWER

OBSERVED DISTRIBUTION OF RESPONSES

27	49	19	5

RIGHT ANSWER

$$X^2 = \frac{(27-17)^2}{17} + \frac{(19-17)^2}{17} + \frac{(5-17)^2}{17}$$

$$= \frac{100 + 4 + 144}{17}$$

$$= \frac{248}{17}$$

$$= 14.59$$

$p \leq .01$

OBSERVATION 2

ANSWER DISTRIBUTION

ITEM	STUDENT A	STUDENT B
1	✓	✓
2	✓	✓
3	✓	A
4	C	✓
5	D	D
6	✓	✓
7	B	B
8	B	✓
9	✓	C
10	D	A
SCORE	5	5

RIGHT ANSWERS IN COMMON: 3

WRONG ANSWERS IN COMMON: 2

BACKGROUND OBSERVATIONS

In 1958, I took a graduate summer school course in test construction. The procedure for items analysis we used involved making a large chart with an entry for every individual answer given, with an X for wrong answers and a blank for right answers. It struck me that changing the particular wrong answer chosen to X might lose some "information". The first value I found for recording every alternative selected was as a check on the accuracy of my marking. Much later I made two observations which have considerable implication to the present discussions.

Both these observations are illustrated in Figure 0.2. (opposite)

First, if wrong answers are chosen blindly, each wrong alternative should be chosen about equally. The upper left example is the expectation under the KNOW - GUESS assumption. More common is the pattern illustrated as observed. A simple χ^2 calculation using only the wrong answer selections showsthat the observed events are not a statistically random event. Our concern here is - "are wrong answers blind guesses?" The observation recorded suggests that this question should be answered, "No!"

The second observation that I made is illustrated on the right in Figure 0.2. Here we have a simulation of two students, both of whom have 5 items out of 10 correct. The performance of these two are usually considered to be equivalent. However, these two students have only 3 right answers in common. It seemed to me that the particular answers correct had some meaningful relationship to what each student knew.

Where the wrong answers are concerned, there is a much smaller chance that students have wrong answers in common than right answers in common.

For the right answers, common answers are considered to be systematically selected for a common reason (the students know this material). If the same assumption can be made for wrong answers, i.e. "students may tend to select

STUDY 1.1 POWELL (1968)

PROOF THAT THE "KNOW - GUESS" ASSUMPTION ABOUT WRONG ANSWERS IS FALSE.

INSTRUMENTATION THE PROVERBS TEST GORHAM (1956)

SAMPLE YOUNG ADULTS $N_1 = 18$ $N_2 = 23$

RIGHT ANSWERS

I

COMPREHENSION
29 out of 38*
$S^2 = 76\%$

OTHER 9 ITEMS SHOWED NO CLEAR PATTERN.

WRONG ANSWERS INCLUDES THOSE SELECTED BY > 10% OF GROUPS.

SIMPLIFICATION	ADDITION	SUBSTITUTION	IRRELEVANCY**
5 ITEMS	4 ITEMS	3 ITEMS	3 ITEMS

ANOTHER 4 ITEMS SHOWED NO CLEAR PATTERN.

REPORTED REASONS FOR WRONG ANSWER SELECTION 64% CONSISTENT BETWEEN THE TWO GROUPS.

WRONG ANSWERS ADD 10 % TO VARIANCE.

NOTES

* 2 ITEMS NOT INCLUDED SINCE MORE THAN 90% ANSWERED THEM CORRECTLY.

** NEAR TCP COMP. + HIGH ON IRR. = DIVERGENT THINKERS.

wrong answers which occur in common for a similar systematic reasons," then wrong answers may contain achievement information

It was upon this observational basis that I began (10 years ago) to study the properties of wrong answer selection.

STUDY 1 THE KNOW - GUESS HYPOTHESIS

Do people know the answer or guess blindly? The Proverbs Test (Gorham 1956) proved to be a good test to explore this question. In the first place, it involves translation which in Bloom's Taxonomy (1956) is the second, or Comprehension level. It, therefore, involves more than isolated recall, for which the KNOW - GUESS hypothesis may be valid. The wrong alternatives were selected from among common wrong answers given on the free response version to these same proverbs by patients in mental hospital. For this reason a systematic logical bias among these alternatives should be largely absent.

The test was given to two classes of Junior and Senior College students; who were asked to record their reason for selecting each alternative. Reasons were asked for in order to detect any systematic responses.

A strong single factor was found among the right answers and four factors were found among the wrong answers. (See Figure 1.1, opposite)

The most common reasons in each wrong answer factor was used as a basis for interpreting the factor. Thus the interpretation of each of the four factors was produced with as little inference as possible.

All this analysis was conducted on the smaller of the two groups (N = 18) while the larger group (N = 23) was used for cross-validation.

With all these precautions to minimize Type I error, still 64 per cent of the reasons reported by the larger group were logically equivalent on a factor by factor basis with the reasons reported by the smaller group. These reasons were also behaviorally consistent within each factor.

Thus, it is evident that the hypothesis "All wrong answers on multiple-

PROOF THAT THE "LINEAR DEPENDENCY" ASSUMPTION ABOUT WRONG ANSWERS IS FALSE.

STUDY 2.1 POWELL & ISBISTER (1974) FACTOR PATTERN FOUND VARIMAX ROTATION

N = 307

choice tests are selected randomly," is false. This refutation is empirical not logical. The word "all" in the hypothesis has been struck down by a single contrary case.

People do NOT simply KNOW or GUESS, they use some systematic approach to answering. For example, some high scoring students also made frequent selections in the Irrelevancy class of error for perfectly logical profound reasons. Is this behavior constructively divergent reasoning? For more details about this study see Powell (1968).

We must now dispense with the LINEAR DEPENDENCY problem.

STUDY 2 The LINEAR DEPENDENCY PROBLEM

For this study my research assistant and I developed a complex 30 item higher mental processes multiple choice test. (See: Powell and Isbister - 1974, for a description of the test and findings summarized here).

The mean total correct score on this test was about 13 out of 30 meant many wrong answers to study. Infrequently selected alternatives were eliminated from analysis and some small wrong answer categories (Over Generalization and Over Simplification) were combined.

The results are illustrated in Figure 2.1. If wrong answers were linearly dependent upon total correct scores all wrong answer scores should have appeared at a statistically significant level in the first factor with negative factor loadings (since the total correct appeared with a positive loading.) None of them did so.

There was no evidence for either linear dependencies or structural dependencies which might account for the pattern observed.

Thus, it is now evident that the linear dependencies between right and wrong commonly expected do not occur in higher mental process tests.

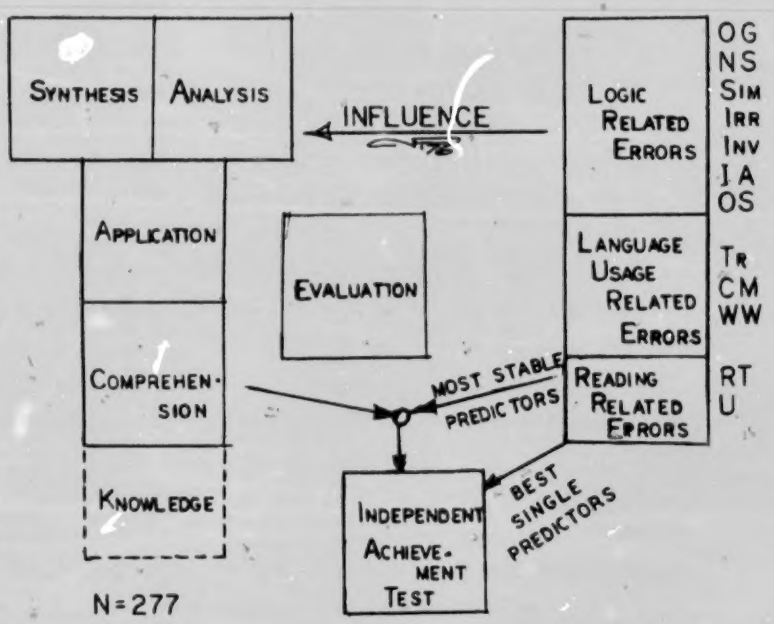
This observation means that the equation: T = R + W + O (Total possible = rights + wrongs + omits)

is both empirically and psychologicall[illegible]. (Page 18, Figure 0.1.2)

STUDY 3.1 PROOF THAT WRONG ANSWER CONTAIN ACHIEVEMENT INFO

RIGHT ANSWER HIERARCHY
AFTER: BLOOM (1956)
FROM: POWELL (1970)
AND MADAUS et al (1973)

WRONG ANSWER HIERARCHY
FROM: POWELL (1970)



GENERAL CONCLUSION:

IT IS THE MANNER IN WHICH A PERSON APPROACHES A QUESTION THAT DETERMINES HOW S/HE ANSWERS IT.

It may be from these observations that the Class R (right answers) and/or the class W (wrong answers) may form generalizations which are too broad to be empirically useful.

It is still necessary to show that wrong answers contain potentially useful information about avhievement.

STUDY 3 ACHIEVEMENT INFORMATION FROM WRONG ANSWERS

In this study (Powell 1970) the experimental test developed for Study 2 (Powell & Isbister - 1974) was used with 277 mature (summer school) adults. They were also given 2 achievement tests in the course one concurrent as the midterm test, the second at the end of the program. The students were randomly assigned to two groups, one for analysis the other for cross-validation.

The most important findings as shown in Figure 3.1. These are: -

1. The order of predictive power was wrong answer subtests > right answer subtests > total correct score.
2. Wrong answers seem to form a hierarchy which influenced the level of the entire item. For an item to be at the analysis level - most or all alternatives had to involve analysis. If wrong alternatives could be eliminated by comprehension strategies, the item became a comprehension item.
3. When right and wrong answers were taken together within the hierarchy the most stable predictions under cross validation resulted.

In general, it seems that:

"It is the manner in which a student approaches a question which determines the answer s/he gives or selects."

Wrong answers may occur because the learner used a different, though not always inappropriate, approach to the solution of the problem. The nature of the answer may reveal the nature of the reasoning involved - giving the teacher

STUDY 4.1
IS THERE A DEVELOPMENTAL SEQUENCE OF WRONG ANSWERS ?

Simplex[a] order of the correlation matrix of relationships among subtests.

N=548

Right Answer scales	Wrong Answer scales													
ABSTRACT		1.00												
	14 Y+[b]	.25	1.00											
	13 Y	.06	.07	1.000										
	12 Y	.00	.01	-.07	1.00									
	Bimodal	-.20	-.01	.04	.05	1.00								
	Alternatives													
	11 Y	.05	-.09	.05	.06	.01	1.00							
	10 Y	-.22	-.16	-.01	.01	.09	.14	1.00						
	9 Y 2	-.25	-.11	-.08	-.16	.02	.06	.17	1.00					
	9 Y 1	-.44	-.18	-.10	-.08	-.04	-.08	.03	.05	1.00				
	8 Y 3	-.46	-.14	-.10	-.05	-.01	-.10	-.01	-.06	.15	1.00			
	8 Y 1	-.48	-.16	-.16	.00	.02	-.24	-.09	.06	.02	.36	1.00		
CONCRETE		-.73	-.22	-.18	-.02	-.04	-.27	-.10	.06	.16	-.14	.37	1.00	
	8 Y 2	-.60	-.15	-.14	-.06	.04	-.24	-.01	.12	.23	.26	.26	.39	1.00
	8 Y 4	-.31	-.16	-.05	-.04	-.05	-.07	-.08	-.02	.04	.06	.22	.21	.16	1.00
		ABS	14 Y+	13 Y	12 Y	Bimodal	11 Y	10 Y	9 Y 2	9 Y 1	8 Y 3	8 Y 1	CON	8 Y 2	8 Y 4

NOTES: a. The Simplex arrangement involves ordering the correlations in such a manner that the magnitude of the correlations increase vertically upward and horizontally to the right.

b. In order to have a reasonable number of children in this classification 14, 15, and 16 year olds were combined.

YES

a basis for corrective action.

The study used rectilinear analytic procedures (factor analysis and stepwise multiple-linear regression), but the researcher felt that these weak results might be strengthened by non-linear procedures.

The main criticism of my Dissertation was that magitude of my statistical results were not large enough nor the cross-validations stable enough to be certain that the results were not statistical artifacts.

The question "would these observations replicate?" was restated in Study 4 as "Is there a developmental sequence of cognition related to wrong answers?"

STUDY 4 A DEVELOPMENTAL SEQUENCE OF WRONG ANSWERS

Study 4 used the Proverbs Test in a near replication of Study 1 with respect to method except that the test was given to 548 children in Grades 3 to 8 inclusive. These children were interviewed for their reasons by trained interviewers. No cross-validating group was used. This study also had a predictive component similar to Study 3.

It was hoped because of a concrete right answer scale that the test could be used to examine the transition from concrete to abstract thinking in children.

For the full details of this report see Powell (1976b).

Figure 4.1 shows the simplex arrangement of the correlation matrix among the subtests used (both right and wrong). The simplex arrangement can be used to suggest the scale properties of a set of variables. The code used in this Figure is age related such that 8 Y 1 means the first of 4 arbitrarily numbered wrong answer subtests related to 8 year olds.

Note that there is not a single deviation from the actual chronological progression of the wrong answer subtests in this arrangement. This observation leaves little doubt that the orderly progression of these subtests is age

STUDY 4.2
EXAMPLE INTERPRETATION

PROVERB:

QUICKLY COME, QUICKLY GO.

TRANSLATIONS:

1. ALWAYS COMING AND GOING AND NEVER SATISFIED.

CHARACTERISTIC OF 13 YEAR OLDS.

"YOU SHOULD STICK TO A JOB TIL IT'S FINISHED."

2. WHAT YOU GET EASILY DOES NOT MEAN MUCH TO YOU.

CHARACTERISTIC OF ADULTS.

3. ALWAYS DO THINGS ON TIME.

CHARACTERISTIC OF 8 YEAR OLDS.

"THAT'S WHAT TEACHER ALWAYS SAYS."

4. MOST PEOPLE DO AS THEY PLEASE AND GO AS THEY PLEASE.

CHARACTERISTIC OF 10 YEAR OLDS.

"IT TALKS ABOUT COMING AND GOING."

dependent. That is, there is, in fact a developmental sequence among wrong answers.

Figure 4.2 gives an illustration of the reasoning reported by the children in this age progression based upon one item. The reasons show a steady progression in three distinct stages from self-contained personal experience through using language literally (i.e. in a denotive sense) to effective use of language in a figurative (connotive) sense. Full delates (from Powell 1976b) of the interpretation of each subtest are given on Table I.

The procedure used to establish these subtests was:

1) determine which particular wrong answers were most characteristic of which age level.
2) correlating these age characteristic answers to establish homogeneous groupings.
3) Selection only those groupings with 4 or more members as subtests.
4) Scoring these subtests in the usual manner, with equal weight for each member.

This procedure produced 12 subtests which, when arranged in the simplex order, seemed to progress by age from extreme over-reduction of the data to integration of the data in a cyclical pattern which occurred at each developmental stage.

At this point we have two separate process defined sequences of wrong answers from two different studies with two different tests given to two different populations. These are Studies 3 and 4. Since there are some striking similarities among some of the descriptions it is reasonable to examine the similarity between these two sequences. Figure 4.3 (Page 35) gives these similarities and reports Spearman's rank order correlation coefficient (Rho) between descriptively equivalent classes. Since the coefficient is so large (.86) there is little question about the presence of an orderly

TABLE I

STUDY 4.3
CONSISTENCY FOUND
BETWEEN STUDY 3 AND STUDY 4

$\rho = .86$

STUDY 5.1
TYPICAL PATTERN OF ALL ANSWERS ON AN ITEM CROSS-TABULATED WITH THE AGE OF THE RESPONDANTS

RESCALED TO REMOVE GROUP EFFECTS

ITEM X ALTERNATIVE

AGE	A	B	C	D	N
8	13	19	57*	11	100
9	17	25	46	12	100
10	25	28	31	16	100
11	31*	32	20	17	100
12	21	44	15	20	100
13	16	52*	11	21	100
14+	14	49	9	28*	100

RIGHT ANSWER

MODAL VALUE

relationship among these events. This observation suggests that the hierarchy found in Study 3 is genuine and not a statistical artifact.

Evidently, children at different age levels characteristically interpret the same problem quite differently, in a similar manner to adults who are functioning at different levels in the hierarchy.

So much for the psychological aspects of wrong answer interpretation. There appears to be a genuine, empirically based phenomenon among these data which has not been here-to-fore observed. We must now address the statistical issues raised in this series of studies.

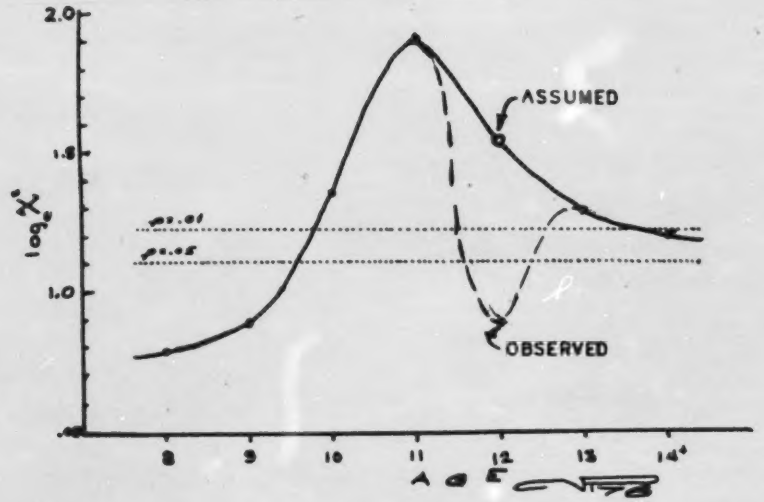
STUDY 5 THE STATISTICAL PROBLEMS ARISING FROM WRONG ANSWER ANALYSIS

Figure 5.1 illustrates how the age characteristic wrong answers were determined.

The four responses for each item were cross-tabulated against the age in years of the respondent. It was noted immediately that each alternative had its own pattern. The right answers, for instance tended to increase steadily with age. The fact that, for this item, the 14+ year olds selected the right answer slightly less frequently can be accounted for quite easily. The 14, 15 and 16 year olds included in this group represent the children who have been delayed in their transfer to secondary school as a result of slower than average progress.

One wrong answer class (alternative C) has its modal (highest frequency) selection with the 8 year olds as would be expected based upon the linear dependency assumption. This frequency is seen to decline from the 8 to the 9 year old group. But it declines more rapidly (11 people) than the right answering group increases (6 people). The remainder split between the other two wrong answers.

STUDY 5.2
OBSERVED CURVE
FOR THE 11Y
CLUSTER



NOTE: Part of the 12 year old group appears to be atypical. The assumed value plotted is the geometric mean between the 11 year olds' χ^2 and the 13 year olds' χ^2 values.

CALL THIS A "PHASE."

This observation implies that the developmental sequence may not involve immediately getting the answer right. Instead the child may shift to a higher order error.

The two remarkable observations are; 1) alternative A where the frequency of selection increases to a high point at the age of 11 years and then declines and 2) alternative D where there is a steady increase with age.

When all 160 alternatives were considered, all age levels showed modal peaks for particular wrong answers. Also a group of a dozen appeared as bimodal. Those alternatives with modes in the same age level were put together, and the bimodal was treated separately.

Right answers, appeared largely homogeneous with the concrete right answers having practically all their modes at the age levels of 8 and 9, and the abstract answers having practically all their modes at the age levels 12, 13 and 14, with most of them at 13.

Figure 5.2 gives an example for the 11 year subtest cosmetically adjusted by using $\log_e \chi^2$ to reduce the kurtosis (peakedness) of the curve.

The emergence and disappearance of a single behavior such as illustrated in this figure can be used to supply a statistically based definition of a phase in development. This definition would seem to be in agreement with the minor qualitative changes which Piaget calls phases as well.

Intercorrelations among age characteristic alternatives ultimately produced 12 wrong answer subtests and led to the dropping of 9 alternatives which did not meet the homogeneity criteria. Hence the procedure reported here accounts for 151 of the possible 160 alternatives on the test.

TABLE II

ABSTRACT RIGHT ANSWER
DISTRIBUTION AND THE CALCULATION
OF SCALE "A" AND SCALE "P"

GROUP	EXPECTED (X 100)	RANGE INCLUDED	8	8/100	9	9/100	10	10/100	11	11/100	12	12/100	13	13/100	14+	14+/100
20	1.0-10	39-40							1	1						
19	1.0-8	37-38							1	1	2	3	2	3		
18	2.0-7	35-36					3	4		-	4	5	5	6	7	13
17	3.0-6	33-34			1	1	3	4	6	3	6	8	7	9	2	4
16	4.0-5	31-32			2	2	2	2	5	6	4	5	6	8	3	6
15	3.0-4	29-30							5	5	5	6	14	18	6	12
14	3.0-3	27-28							15	15	10	13	7	9	9	17
13	.02	25-26					3	4	3	3	4	5	15	19	6	12
12	.08	23-24			1	1	3	4	5	6	5	6	6	8	3	6
11	.3	21-22	2	4	2	2	2	2	15	15	10	13	14	18	6	12
10	1.0	19-20	1	2	4	4	7	8	8	8	11	14	15	19	6	12
9	2.7	17-18	-	-	3	3	9	11	7	7	8	11	6	8	3	6
8	6.1	15-16	1	2	9	9	13	15	21	20	7	9	6	8	6	11
7	11.2	13-14	5	9	7	7	8	10	10	10	7	9	4	5	6	11
6	16.9	11-12	6	11	13	13	8	10	6	6	5	6	3	4	1	2
5	20.2	9-10	7	12	20	19	11	13	8	8	6	8	1	2	1	2
4	19.0	7-8	10	19	18	18	12	14	5	5	1	2	-	-	1	2
3	13.4	5-6	11	20	20	18	5	6	4	4	-	-	1	1	1	2
2	6.7	3-4	9	17	5	5	2	2	1	1	1	1				
1	2.1	1-2	2	4	1	1	1	1								
0	.3	0	-	-												
			54	100	101	100	84	100	101	100	77	100	77	100	52	
	χ^2			79.4		36.1		1096.3		364260.4		332096.9		379818.6		3313.7
	P[a]			-1.5		-1.1		+2.8		+5.7		+5.7		+5.7		+5.9
	A[b]			.22		.27		.34		.63		.48		.54		.52

NOTES: a. Scale "P" is found from the χ^2 values as follows: for positive values $P = \frac{1}{4}(\log_e \chi^2_o - \log_e \chi^2_{.05}) + 1$ where χ^2_o is the obtained value and $\chi^2_{.05}$ is the value for P = .05 Negative values of P (which are used when the distribution of observations tends below chance level) are calculated the same way but carry a negative sign when reported.

b. Scale "A" is the average proportion selected thus: $A = \frac{\sum X_i}{Nn}$ where X_i is the ith individual's score on the subtest. N is the number of members in the group and n is the number of items in the subtest.

c. Overall P=20X (A - .25) (Actual for this sample 20.2 times).

The simplex analysis reported in Figure 4.1 reports the developmental sequence found. The subtests were interpreted using the reported reasons for selection to produce the results reported in Table I (Powell 1976b). Between 52.9 per cent and 62.7 per cent consistency was found within each subtest, however, children showed a relatively high level of non response.

The pattern for each subtest was then calculated using two procedures. The first procedure, or "A" scale (for average) which represents the selection ratio for the subtest. The second procedure, "P" scale (for Powell) is my own development. It is based upon the χ^2 value using the binomial probabilities for the expected values. Several cosmetic adjustments to χ^2 were made as follows. 1) taking the square root, 2) taking the natural logarithm, and 3) rescaling against the natural logarithm for the .05 for χ^2 for that subtest. Negative values for P were produced in the same manner except that the curve was shifted below the expectation pattern. Absolute values less than ±1.0 mean that the χ^2 were not significantly different from chance. Coincidentally these adjustments produced an overall scale value for P which very nearly equal to 20 times the value of A - .25.

A detailed example of this procedure using the Abstract scale is found in Table II.

This procedure can be used to give a general picture of the trends of a particular subtest across time but not the interpretation of specific scores. The maximum observed or maximum possible score can be age referenced to the mode, but all other values are ambiguous because they have two possible values on the age axis for each value on the score axis.

This problem does not exist for straight line procedures. It, therefore, becomes necessary to determine whether straight line or curved line functions best describe these observations.

TABLE III

STEPWISE MULTIPLE REGRESSION ANALYSIS
($p \leq .100$)

CRITERIA (GATES-MACGINITIE)	GRADE 4 PREDICTORS (PROVERBS)	R^2	p=	GRADE 7 PREDICTORS	R^2	p=
LINEAR PREDICTIONS						
1. Speed & Accuracy						
a. correct	11y	.109	.006	10Y(-)	.037	.068
	8Y3(-)*		.055			
b. attempted	CON(-)	.032	.075	—	—	—
c. wrong	12y	.091	.010	1RR(-)	.097	.029
(b - a)	8Y2		.024	8Y1		.062
2. Vocabulary	—	—	—	1RR(-)	.061	.019
3. Comprehension	11y	.057	.017	14Y	.107	.010
						.072
4. Composite (2 + 3)	9Y1	.044	.035	11Y(-)	.109	.005
				10Y		.064
QUADRATIC PREDICTIONS						
1. Speed & Accuracy						
a. correct	11Y	.332	.000	ABS	.231	.000
	ABS		.005			
	9Y1		.011			
	9Y2(-)		.028			
	8Y4		.042			
b. attempted	11Y	.271	.013	ABS	.124	.004
	ABS		.002	CON		.020
	9Y1		.017			
	9Y2(-)		.028			
	8Y4		.038			
c. wrong	12Y	.362	.057	14Y	.255	.000
(b - a)				10Y(-)		.011
				CON		.077
2. Vocabulary	ABS	.431	.001	13Y	.389	.000
	9Y2(-)		.000	ABS		.011
	8Y2		.001	14Y(-)		.077
	14Y		.026			
	8Y3(-)		.049			
3. Comprehension	8Y4	.413	.017	ABS	.301	.000
	8Y2(-)		.001			
	ABS		.001			
	9Y2(-)		.001			
4. Composite	8Y2(-)	.459	.001	13Y	.407	.000
	ABS		.000	ABS		.067
	9Y2(-)		.001			
	8Y4		.044			
	14Y		.045			

Figure 5.3 summarizes these results. Six scores on a standardized reading test (Gates - MacGinitie) were available for the Grade 4's and the Grade 7's. The 14 subscores were used to predict each of these achievement scales. Both rectilinear (straight line) and quadratic (parabolic) predictions were used. In the rectilinear functions, a significant positive relationship means that the trend of the subtest is upward to the right as the scores on the reading test progress in the same direction. Negative values indicate downward trends. For the Grade 4 group the 11Y scores show this positive trend and the concrete scores show a negative trend. Non-significant values for straight line predictions can arise from horizontal trend lines or from variables unrelated to the scale being predicted. (eg. B/M)

Where the quadratic prediction equations are concerned (using Grade 4 for the examples) an upward turned cupshaped curve will be positive, (eg. ABS) downward - negative, (eg. 9Y2) and a straight line (eg. 11Y) or an "S" shaped curve (eg. CON) will appear as not significant in the predictions. Using an arbitrary definition of an inclination less that $\pm$.5 P as horizontal the prediction results are compared with the exposed portions of the observed trend lines for these 14 variables. Three observations are important. First wrong answers are generally better predictors than right answers. This replicates the findings in Study 3, suggesting that that finding was probably not a statistical artifact either.

Second the quadric predictions emerged as substancially better predictors than the straight line predictors, by a factor on the average of at least 4 (.07 to .28 for Grade 7). Third, the shapes and inclinations of the trend line patterns are supported in at least 68 per cent of the cases (19/28 for Grade 7).

Table III gives the details upon which these observations are made.

Figure 5.3 and Table III face each other on pages 42 and 43.

It appears, then, the problem of the ambiguous interpretation of "phase"

trend lines may be unavoidable because curvilinear trends enjoy a fair amount of empirical support from these data.

When these trend lines are reproduced in their entirety the results appear as shown in Figure 5.4. These curves are the hand fitted averages of the "A" and "P" scales. Table IV gives the data base from which these curves are derived. Figure 5.4 and Table IV are found facing each other as pages 46 and 47.

The band between $\pm$1 is not significantly different from chance on the P scale. Since P$\approx$20x(A -.25), the band between .20 and .30 on the A scale must also be within this accidental range. Only the Bimodal (B/M) Irrelevancy subtest remains with this band throughout its entire length. All other curves have at least part of their length outside of this range. Five of the 14 curves cross from significantly above to significantly below expectation by chance alone. Three of the curves cross in the opposite direct, and two curves go from significantly below to significantly above and back to significantly below again. There is no question in these latter two cases that a curvilinear interpretation must apply. The improvement gained by using curvilinear predictions and the degree of agreement with the hand fitted curves lends strong support to the possibility that curvilinear trend lines generally apply. These data are further supported by the observation I made in my Dissertation (Powell 1970) that the further I departed from rectilinear mathematical analysis the more meaningful the results became.

When a path analysis is performed upon these data the results appear as in Figure 5.5 (on page 48). In this case a linear dependency clearly emerges - between concrete right answers and abstract right answers. This pattern has psychological as well as mathematical validity. This one variable

TABLE IV

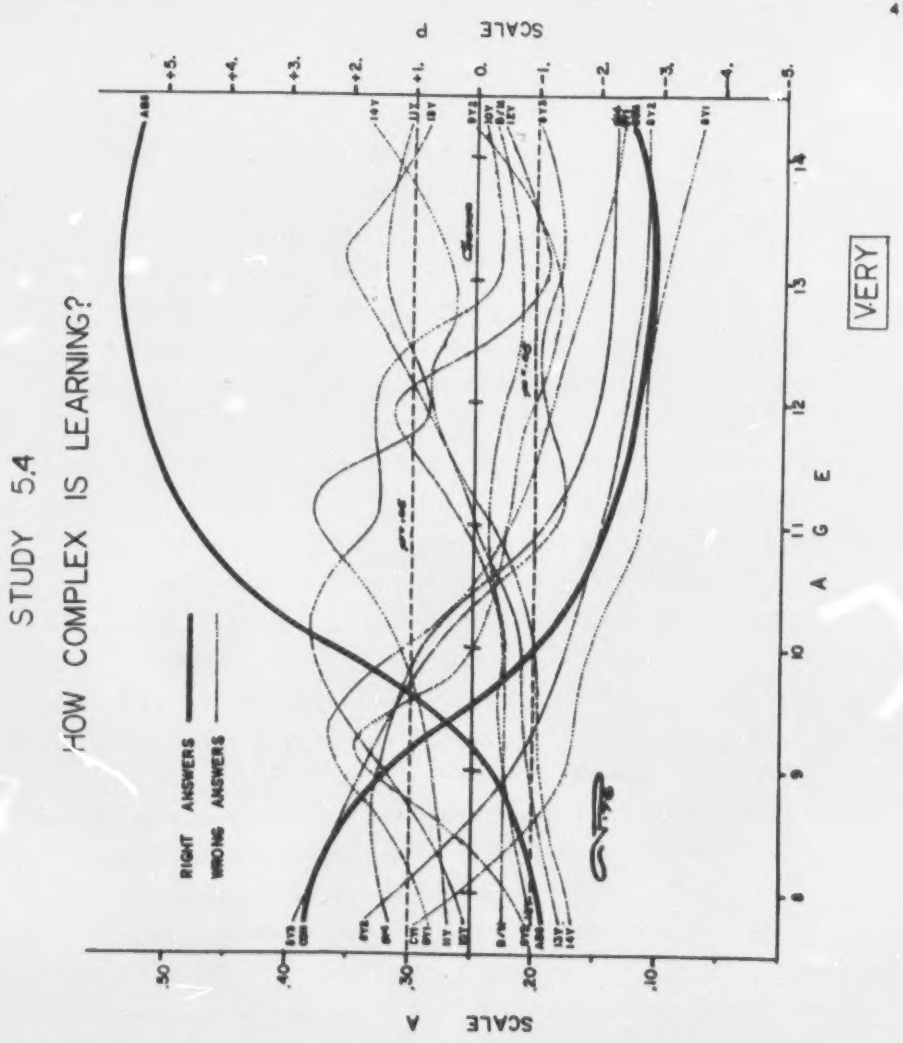
Proportional Subscores
by age on each subtest

AGE GROUPINGS

Sub-test	Max. possible[5] score	Scale A[1] 8	9	10	11	12	13	14+	Scale p[2] 8	9	10	11	12	13	14+
CON[3]	17+3	.38	.27	.19	.14	.11	.08	.09	2.6	2.3	-1.2	-1.8	-2.2	-2.6	-2.3
6Y4	5	.32	.32	.28	.19	.14	.13	.12	1.4	1.5	.8	-1.3	-1.8	-1.9	-1.8
8Y3	5+2	.39	.34	.28	.18	.19	.16	.19	2.3	1.6	.5	-1.2	-1.2	-1.6	-1.1
8Y2	12+3	.30	.23	.19	.14	.12	.11	.11	1.6	-1.0	-1.4	-2.1	-2.2	-2.2	-2.2
8Y1	8+1	.26	.16	.15	.10	.10	.07	.04	.3	-1.2	-.17	-2.1	-2.3	-2.4	-2.6
9Y2	5	.22	.33	.24	.24	.20	.17	.25	-.5	1.7	-. 2	-1.0	-1.0	-1.4	- .4
9Y1	8	.27	.36	.30	.22	.18	.14	.12	.5	2.0	1.1	- .8	-1.2	-1.6	-1.9
10Y	8+1	.26	.32	.39	.31	.27	.24	.25	.2	1.5	2.1	1.2	1.4	- .7	- .4
11Y	6	.26	.29	.30	.39	.28	.33	.32	.6	.7	1.2	1.9	.7	1.3	1.1
12Y	4	.19	.22	.26	.26	.34	.19	.23	-.9	-.7	-.9	.2	1.2	-1.2	- .9
13Y	6	.19	.20	.21	.25	.27	.36	.29	-1.1	-1.1	-.7	.3	.8	1.8	1.0
14Y	7	.16	.19	.18	.24	.28	.25	.32	-1.5	-1.0	-.8	-.6	1.1	.4	1.3
Bimodal[4]	10	.24	.24	.23	.24	.23	.21	.23	- .6	-.6	-.3	-.5	- .6	-.8	-.2
ABS[3]	40	.22	.27	.34	.43	.48	.54	.52	-1.5	-1.1	2.8	5.7	5.7	5.7	+5.9
Items Used Numbers in each group	151	54	101	84	103	77	77	52							
Not used	9														
Total	160														

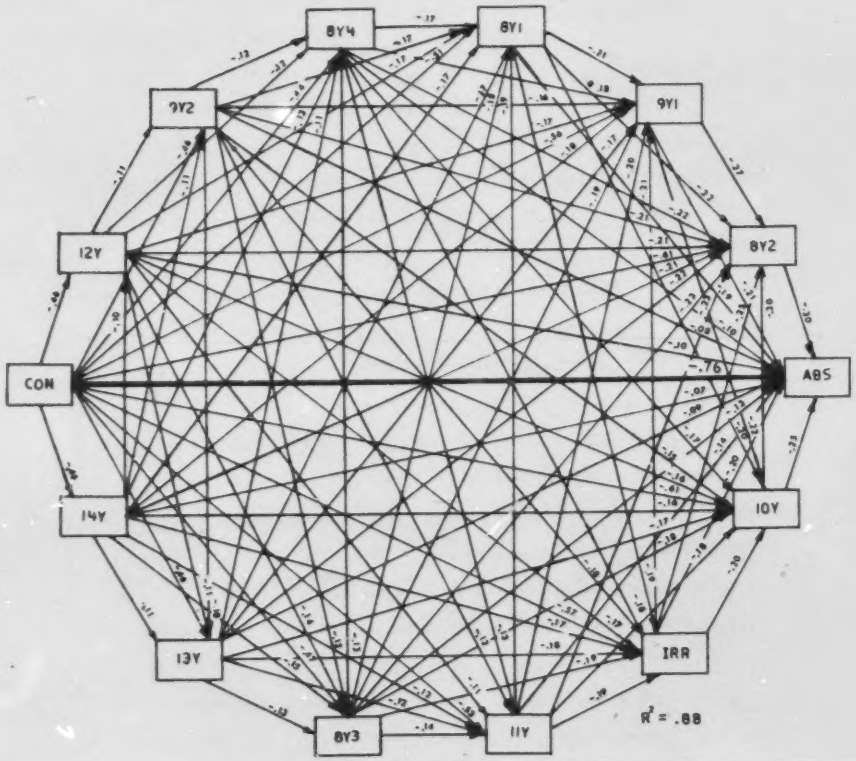
NOTES:

1. Scale A is the average proportion selected in each subtest.
2. Scale P is adapted from the χ^2 value such that when the displacement tends to be upward then $P = \frac{1}{4}(10g_e \chi^2_a - \log_e \chi^2.05) + 1$.
   where χ^2_a is the obtained value and $\chi^2.05$ is the value of χ^2 needed for $p = .05$.
3. Right answer subtests are given in capital letters.
4. Items in this set were clearly bimodal by age selection.
5. In some subtests more than one alternative is used in one or more particular items. In this case the convention 17+3 is used to indicate that the subtest involves 17 items but 20 alternatives are scored in the subtest.



STUDY 5.5

WHAT DOES THE PROVERBS TEST REALLY MEASURE?



THE TRANSITION FROM CONCRETE TO ABSTRACT LANGUAGE USAGE

accounts for 58 per cent of the variance $(-.76)^2$. The strongest secondary relationships are between concrete right answers and wrong answers. (The relevant standardized β weights are reported near the head of each arrow.) This supports the observation made earlier (See: 5.1) that at least part of the transition from concrete to abstract language interpretations involves a transition through wrong answers.

Further details are more difficult to conclude because this was a cross-sectional study. However, concrete thinking seems to be most strongly followed by correct answers related to literal interpretation (10Y) and to the breakdown characteristic of a new stage (Isolated Responses 8Y2). In third place is the bimodal (Irrelevancy) subtest which has already been found (in Studies 1 and 2) to be correct answers when chosen by a high scoring group. Beyond these observations there is insufficient data to establish the nature of the typical transformations which occur with development.

On the other hand it is very clear that,

1. The path from concrete to abstract thinking is not necessarily direct.
2. Wrong answers are not linearly dependent with right answers on this test. The 12 subtests together account for only about 30 per cent of the variance.
3. The test has a high structural consistency with 88 per cent of the variance accounted for by the path analysis.
4. The test should be considered a valid instrument for studying the transition from concrete to abstract language usage.

The evidence, thus far, supports a phase model for development, but what about stages in this sequence? This possibility is more difficult to produce. However, visual examination of the trend lines in Figure 5.4. suggest

STUDY 5.6

EVIDENCE FOR A STAGE PATTERN ?

NUMBER OF CURVES IN A TREND. (FROM 5.4)

TREND*	AGE BELOW 8?	8-9	9-10	10-11	11-12	12-13	13-14+
UP	3	5	2	4	4	2	3
LEVEL	6	4	6	4	7	7	7
DOWN	3	3	4	4	2	3	2

a general upward trend among wrong answers between ages 8 and 9, 10 and 11 and 13 and 14. It would appear that downward trends or at least balanced results occur, below 8 years, 9 to 10, 10 to 11, and 12 to 13. The upward trends seem to follow immediately after the emergence of a new stage as derived from reasons for answers, except, of course, for the 13 to 14 increase which has already been explained. These data are summarized in Figure 5.6.

It would seem that overall error frequency may increase with a major transition. This observation is in agreement with Piaget's speculations. He suggests that with each major accommodation, the person's total schemata must be restructured. Fragmented responses such as Isolated Responses, Word Associations, and Over Simplifications may be the first indication of entrance to a new stage. These may be followed by progressive improvement until the stage is consolidated, at which point a new stage is possible.

It must be remembered, however, that our present schools tend to reward convergent behavior and to punish or at least not reinforce divergent behavior including errors. Is it any wonder, then, that 25 per cent of the 13 year old children in this sample were still functioning in the personalising stage? Could it be that the behavior modification concept of errorless learning is itself a fallacy?

In any case both observational and mathematical support for a phase and stage pattern of human development have been obtained from the pattern analysis of wrong answers.

It would appear that learning is vastly more complex than a sequence of total correct scores would seem to imply. Also, since predictions from wrong answers and non-linear (or curvilinear) approaches appear to be more empirically valid than linear approaches, the use of total correct scores may not be valid. The implications to education, of such a suggestion are indeed profound.

CONCLUSIONS AND IMPLICATIONS

The general picture of the outcomes of research is the area of human resources development is given in the sketch opposite.

The study of convergent behavior is upon and active within the context of behavior modification. Contingency management analysis is the key to this research. The appropriate mathematical procedures would seem to be rectilinear, and the thrust would seem to be normalizing or "making everyone the same".

Where divergent behaviors are concerned, very little research has been done by comparison. The series of studies reported here suggest that response pattern analysis is the key, that nonlinear or at least curvilinear analytic techniques are necessary, and the thrust would seem to be related to differential development rather than convergence. Hence the "exceptionalization" or "making everyone different" approach to education would seem to be more appropriate. That is, the approach involves enhancing rather than inhibiting individuality.

This latter procedure creates a philosophical problem in that equality cannot be defined in terms of sameness in a society stressing individuality. An alternative definition for equality is needed. If this definition is derived from the leveling effect of the exchange of services, then a third type of behavior the communicative—interactive behaviors need development. Little is known about this area of human behavior, although some progress has been made. For instance, group learning activities, where effective interaction is necessary, produce more cooperation than individual learning activities.

For these three human performance outcomes to be appropriately integrated, there is need for at least some individuals as teachers to be able to go beyond these three. That is, we need self-actualizing individuals. Very little is known about the process of going beyond self. Although we do have

some understanding of the characteristics of people who behave this way.

In any case, the behavior modification approach is probably necessary but is clearly not sufficient to describe human performance. It is probable that divergent behavior analysis is not sufficient either.

THE EVIDENCE

The evidence supporting the proposition that the study of divergent behavior is also necessary has been presented in this report and can be summarized as follows:-

1. The know-guess hypothesis is false.
2. The linear dependency hypothesis is false.
3. Wrong answers are better predictors of achievement than right answers.
4. Wrong answers form a hierarchy which influences item performance.
5. Right and wrong answers combined, a) help identify constructive divergent thinkers, b) provide the most stable predictions under cross-validation conditions.
6. Wrong answers form a developmental sequence.
7. Reasons for selecting wrong answers are reasonable and relate to the approach to the problem used by the learner for both children and adults.
8. Hierarchies of wrong answers for children and adults are closely similar (ρ = .86).
9. Specific behaviors emerge and disappear forming a sequence of phases.
10. Learners shift to wrong answers of higher order as well as to right answers as they develop.
11. Range of error types increases with the transition to each new stage.

CONCLUSIONS

This evidence leads logically to the following conclusions:

1. Both phases and stages are supported by answer pattern analysis.
2. Right answers alone are not a sufficient set to adequately describe development.
3. Answering patterns are apparently not rectilinear.

4. Sufficient evidence now exists to support the contention that wrong answers contain useful achievement information related to information processing as distinct from information recall. Much of this evidence has withstood replication.
5. The characteristic typologies of development transformations with reference to divergent behavior are not yet known.

The greatest current need is for effective practical problem solvers under the constraints of real time and substantial uncertainty. The information explosion has produced a surfeit of information but most of it is not in the appropriate form to be useful in the solution of situationally dependent problems.

WHERE WE ONCE HAD QUESTIONS LOOKING FOR ANSWERS--WE NOW HAVE ANSWERS LOOKING FOR QUESTIONS.

IMPLICATIONS

1. To Education

Typically, the teacher who focuses upon right answers (convergent behavior) ignores or punishes wrong answers. If errorless learning were possible, this behavior would not have serious consequences. The problem is two-fold. First, the evidence of these studies suggests that errorless learning does not occur. Second, the findings of Brophy and Evertson (1976) suggest that even for the optimal development of convergent (right answer behavior) a success rate at about the 80 per cent (not 100 per cent) right answer level is best.

The present findings suggest that particular errors tend to increase at the onset of each phase of development. There is also evidence that several types of error tend to increase with the onset of a new stage of development. Suppose that the learner is externally oriented for rewards, as behavior modification encourages. What would happen to such a child when necessary developmental errors are either ignored or punished? The

findings of behavior modification would suggest that in this situation, learning in the general sense would be expected to cease.

There is evidence in the present studies and from other sources in our work that 25 per cent of the children who transfer from Grade 8 (Grade school) to Grade 9 (High school) are still functioning at the personalizing (8 year old level. (See: Powell, 1976a) We do not yet know whether this failure to progress is a factor of learning capacity or of failure of the schools by using the wrong behavior models.

Another way of stating this same problem is raised by such writers as Hoffmann (1962) who suggests in his book The Tyrrany of Testing that forcing convergence in schools destroys creativity and prevents profound understanding. The findings of the present studies support the implications of this attack upon education.

On the one hand the teacher who is concerned only with what is "right" needs pay little attention to a child other than to match his or her answer with the one the teacher expects. Alternatively the teacher who is trying to figure out why a child gave an unexpected answer must pay a great deal of attention to that child. Also as the teacher and child explore the basis for the answer; the child becomes conscious of the way his or her viewpoint influences his or her problem solving results. This latter procedure can be expected to develop a sense of personal mastery and intrinsic motivation in the same manner that developing conscious awareness of internal states can develop internal mastery in the biofeedback setting.

Teachers who have successfully developed the central aspect of this "whole child" approach to education seem to have produced spectacular learning results. Does divergent behavior analysis actually unlock these powerful motivating forces? Behavior modification has much success in the short run but seems to fail in the long haul. How well might this alternative approach work?

Within the analogies just presented, the implication would seem to be that a response pattern analysis approach may be more powerful than behavior modification in its long term effects.

2. To Research

Response pattern analysis focusses upon differences among behaviors rather than similarities among them. Educational research has failed to demonstrate substantial differential effects with different educational procedures. However, total correct score procedures count across and ignore item by item differences or alternative by alternative differences in performance. Such internal differences have been treated as irrelevant.

It is quite possible that response pattern analysis will expose differential effects not revealed in a total correct score. If this possibility is supported by further observations, large quantities of research which have used rectilinear (convergent) analysis might benefit from reworking. The apparent invalidity of total correct scores throws all of these studies into the questionable validity category.

3. To Test Theory and Analysis

If rectilinear analysis is inappropriate, the whole of test theory needs reworking. It may be that the appropriate mathematical procedures already exist. In this case, extensive research will be needed to determine which procedures fit which situations. If these do not exist, as Sockloff (1976) suggests, then a completely new formulation will need to be generated by applied mathematicians.

Pattern analysis may also need to be applied to all existing standardized tests so that the norms may be reworked to accommodate the interpretation of all answers and not just the right answers. In doing this analysis, the problem of the ambiguous interpretation of curvilinear trend lines will need to be solved.

4. To Learning Theory

Are there a limited number of strategies which children tend to use over

and over? Do they try out their entire reportoire over again with each new stage until they learn when each works and when it does not? This pattern is very similar to the one described by Piaget (1952) in his discussions concerning the relationbetween "assimilation" and "accommodation".

If learning occurs in this pattern, then the observations reported here are explained. This explanation suggests that process transcends product in learning. We would, in this case, be left with many unanswered questions. What are the processes used? In what order do they occur? Which ones are critical in which kinds of settings? What are the effects of specific teaching interventions on process outcomes? How are strategies acquired? Why do many children fail to develop particular strategies? Can a knowledge of divergence be used to force convergence in new and more insidious ways, or does the procedure open a new defence against psychological coersion?

If response pattern analysis is the key to the study of divergent behavior, then many of these questions might be answered using this analytical system.

Is it possible that findings presented here are similar to the child's discovery that words can have connotative meanings? Do these findings open the vast possibilities which the implications drawn seem to suggest? At least the findings are now substantial enough to warrant an extensive exploration of the approach. Perhaps, in addition to coming to grips with some interesting issues, we will deal effectively with a number of criticisms now being levied against our schools, and in consequence meet some of the emerging needs of our post industrial society.

Those who have attempted the test sample at the beginning of this report may wish to analyze their results using the profile data following the References on Page 62.

THANK YOU
FOR
YOUR ATTENTION

YOU TOO
ARE NOW
A
WRONG ANSWER EXPERT!

References

BLOOM, Benjamin S. Taxonomy of Educational Objectives; Handbook I: Cognitive domain. U.V., David MacKay, 1956.

BROPHY, Jere E. and EVERTSON, Carolyn M. Learning from Teaching: A Developmental Perspective. Boston, Allyn and Bacon, 1976. (XV+212 PP).

CARLTON, R. A. Popular Images of the School. Alberta Journal of Educational Research, 20(1), March 1974, 59 - 74.

GATES, Arthur I. and MACGINITIE, Walter H. Gates - MacGinitie Reading Test Teachers College, Columbia, 1969.

GETZELS, J. W. Images of the classroom and visions of the learner. School Review, 82(4) August 1974, 527 - 540.

GORHAM, Donald R. Proverbs Test. Psychological Test Specialists, 1956.

HOFFMANN, B. The Tyrrany of Testing. New York, Collier, 1962.

JAMISON, Dean, SUPPES, Patrick and WELLS, Stuart, The effectiveness of alternative instructional media: a survey. Review of Educational Research. 44(1), Winter 1974, 1 - 65.

PIAGET, J. The origins of intelligence in children. N.Y., International Universities Press, 1952.

POWELL, J. C. The interpretation of wrong answers from a multiple choice test. Educational and Psychological Measurement, 28(2), Summer 1968, 403 - 412.

POWELL, J. C. Achievement Information from Wrong Answers. (Short Title) Unpublished Ph.D. Dissertation, University of Alberta, May 1970. xiii+240 pp.

POWELL, J. C. Attitudes toward school held by students in an urban secondary school. Unpublished research report, University of Windsor. July 1976a incomplete

POWELL, J. C. The developmental sequence of cognition as revealed by wrong answers. Unpublished research report, University of Windsor, 1976b, 14 pp. Derived from a paper delivered to the Annual Meeting of the Ontario Educational Research Council, December 1975.

POWELL, J. C. and COTTRELL, D. J. Popular images of the schools: revisited. Unpublished research report, University of Windsor, February 1976, 11 pp.

POWELL, J. C. and ISBISTER, Alvin G. A comparison between right and wrong answers on a multiple choice test. Educational and Psychological Measurement, 34(3), Autumn 1974, 49[illegible] - 509.

SOCKLOFF, Alan L. The analysis of nonlinearily via linear regression with polynomial and product variables: an examination. Review of Educational Research 46(2), Spring 1976, 267 - 291).

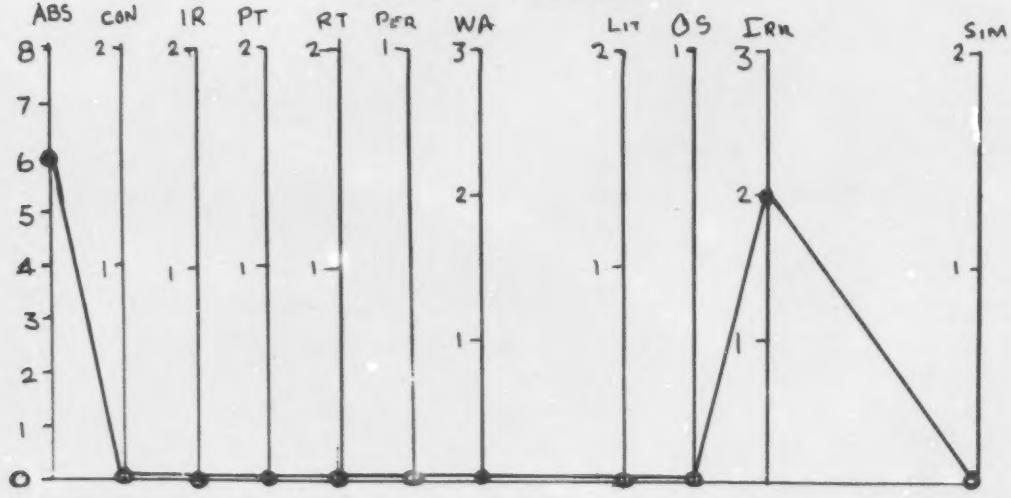
WALKER, Decker F., and SHAFFARZICK, Jon, Comparing Curricula. Review of Educational Research, 44(1), Winter 1974, 83 - 111.

ABRIDGEMENT OF THE PROVERBS TEST

KEY

ITEM	ABSTRACT	CONCRETE	ISOLATED RESPONSES	PARTIAL TRANSLATIONS	REDEFINITION OF TERMS	PERSONALIZED INTERPRETATION	WORD ASSOCIATIONS	LITERAL REDUCTIONS	LITERALIZED INTERPRETATIONS	OVER SIMPLIFICATIONS	IRRELEVANCIES	TRANSPOSITIONS	OVER GENERALIZATIONS	SIMPLIFICATIONS	UNCLASSIFIED
1	C		D				B			A					
2	B	C					D				A				
3	D				B		C		A						
4	C		B								D				A
5	C	B													A, D
6	D								C					B	A
7	C			A		D					B				
8	C			D	B									A	
TOTAL POSSIBLE	8	2	2	2	2	1	3	0	2	1	3	0	0	2	4

PROFILE



TYPICAL "CREATIVE" PERSON